DOCUMENT RESUME

ED 479 922 TM 035 226

AUTHOR	Stapleton, Laura M.; Edmonds, Meaghan
TITLE	An Exploration of the Validity of the Unbounded Write-in Scale.
PUB DATE	2003-00-00
NOTE	23p.
PUB TYPE	Reports - Research (143)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Attitude Measures; Focus Groups; Higher Education; *Rating Scales; *Reliability; Statistical Distributions; *Undergraduate Students; *Validity

ABSTRACT

An exploratory reliability and validity study was conducted of a relatively new response scale developed in the marketing field. Unlike many Likert-type scales, the "unbounded write-in" scale is claimed to produce distributions that more closely approximate normal distributions. This type of scale has been used in large-scale marketing studies. This experimental study sought to determine whether the scale demonstrates item test-retest reliability and whether respondents use the range of the scale in similar ways. (whether two respondents who reported the same level really felt the same way, and whether they used the same point on the scale). Participants were 220 undergraduates who responded to 10 attitudinal items taken from the 1998 General Social Survey and presented as unbounded write-in scales. Results from the study are mixed. Findings suggest that the unbounded write-in scale may offer a reliable alternative to the Likert-type scale, although the claimed advantages of its distributional qualities were not seen in this study. Focus group comments, however, suggest that the scale might not reflect similar attitudes across individuals. Further study appears warranted. Three questionnaire versions are attached. (Contains 3 figures, 4 tables, and 19 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 479 922

An Exploration of the Validity of the Unbounded Write-in Scale

Laura M. Stapleton

University of Texas at Austin

and

Meaghan Edmonds

University of Texas at Austin

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

L. Stapelton

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Correspondence should be directed to the first author at:

504 SZB, Mail stop D5800
University of Texas
Austin, TX 78712
e-mail: *laura.stapleton@mail.utexas.edu*
phone: (512)471-0858
fax: (512)471-1288

TM035226

BEST COPY AVAILABLE

Abstract

This paper shares the results of an exploratory reliability and validity study of a relatively new response scale, developed in the marketing field. Unlike many Likert-type scales, the "unbounded write-in scale" is claimed to produce distributions that more approximate normal distributions and has been used in large-scale marketing studies. However, before its use can be adopted in social science research, it is appropriate to determine whether measurements using this scale are reliable, and equally important, whether the measurements can provide valid representations of attitudes and opinions. This experimental study sought to determine whether the scale demonstrates item test-retest reliability and whether respondents use the range of the scale in similar ways, in other words, whether two respondents who reported the same level really felt the same way and likewise, whether two respondents who felt the same way used the same point on the scale. Results from the study are mixed. Our findings suggest that the unbounded write-in scale may offer a reliable alternative to the Likert-type scale, although the claimed advantages of its distributional qualities were not seen in this study. Focus group comments, however, lead us to believe that the scale might not reflect similar attitudes across individuals. We suggest that our findings warrant further study of the unbounded write-in scale.

An Exploration of the Validity of the Unbounded Write-in Scale

Survey researchers have long been concerned about measurement effects associated with response scales. Among the many response scale issues that have been studied by methodologists is the number of response categories, including how many options are optimum (Masters, 1974; Cox, 1980; Sheatsley, 1983; Peterson, 1985; Smith and Peterson, 1985; Alwin and Krosnick, 1991)). With the advent of computer-based collection of data (either web-based or stand-alone), response scales have become even more flexible, allowing for such scales as the continuous 0-100 bar. These continuous rating scales, however, have been criticized for low reliability and studies have suggested that the lack of anchor points on these scales contribute to unreliability (Parrill, 1999). With response scales in general, Alwin (1992) has found that the greater the number of points on the response scale, the greater the reliability, although there were diminishing returns. Response scale success certainly depends on the context and question stem, however most researchers suggest that seven options, plus or minus two, provide optimum information while maximizing reliability.

Even though the survey field appears to be fairly settled on the choice of number of scale points, the collection of data using five- to nine-point scales often results in non-normal distributions of responses, a condition that violates the assumption of normally-distributed data that exists for many statistical techniques. In attitudinal surveys, researchers often have data which include responses that are clustered at one end of the scale. Statistical methods to handle such non-normally distributed data have been investigated (see, for example, Deshpande, Gore, & Shanubhogue, 1995 and Fouladi, 2000), however, rather than using alternative statistical procedures, it would be preferable to use a response scale that could capture the hypothesized underlying normal distribution.

In the marketing literature, a new response scale option has been introduced that is claimed to capture measurements displaying normal distributions, however the reliability and validity of the scale have yet to be demonstrated. This scale, titled the "unbounded write-in scale," was developed by Eric Marder (1997) and is demonstrated in the example question shown in Figure 1.

Place Figure 1 about here

For this response scale, the respondent is provided with a box, in which he is asked to place as many Ls (to represent liking) or Ds (to represent disliking) as reflects his attitude toward the topic or statement. As suggested by Marder, instead of querying about likes and dislikes of a brand or product, this response scale can be altered to ask about levels of disagreement or agreement with a statement. Such an alteration could lend itself to application in a multitude of social science settings, replacing the common Likert-type questions that ask subjects to respond to statements whether they "strongly agree," "agree," "disagree," or "strongly disagree." In surveys of younger people, in fact, it may be more understandable to the respondent to write in a preferred number of As (for agree) and Ds (for disagree) to graphically reflect their intensity of feeling than to determine whether they believe, for example, "strongly" about something.

Marder (1997) expresses that the unbounded write-in response scale has several attractive features including: 1) it has a natural zero point to represent indifference, 2) it does not require the use of troublesome negative numbers, 3) it is unbounded, so there is no particular ceiling of like and dislike, and 4) it is constructed out of increments of effort that restrain the respondent from indiscriminate excesses and keep the responses within reasonable bounds. A particular advantage with this scale, Marder claims, is that the obtained responses are more normally distributed than responses obtained using Likert-type scales. For example, in a study in which Marder asked two groups of people about

BEST COPY AVAILABLE

their likes and dislikes of fourteen political leaders, comparing the use of the unbounded write-in scale with a +5 to –5 numeric scale, he found that although John F. Kennedy was rated highest of the leaders on average by both groups of people, the distribution of the Kennedy ratings were very different across the two groups. In the fixed numeric scale group, the author reported that the distribution was extremely negatively skewed, with nearly half of the respondents giving Kennedy the top rating of +5. For the unbounded write-in scale group, however, the distribution more closely approximated a bell-curve (although slightly positively-skewed) with 45 percent of the responses having values of L, LL, or LLL.

It is clear from Marder's results that the use of this scale to measure opinions and attitudes offers an intriguing solution to the problem of collecting data that reflect an underlying normal distribution. However, there remain questions of whether attitudes collected on such scales are measured reliably and whether the responses truly reflect the assumed attitudes. Marder has provided no evidence of the reliability of measurements collected with this response scale, only empirical distributions of data collected from large-scale marketing studies. If desirable normal distributions are obtained but the reliability of responses is not assured, then it is doubtful that an improvement in measurement has been gained. In addition, because of the absence of anchor points and the lack of bounds, it is questionable whether the response scale provides valid measurement. It is not clear that LLL to one person indicates the same level of "liking" to another person. An additional concern with the use of the unbounded write-in scale is with the characteristics of the printed box. If the size of the box is related to the number of letters that respondents use, then researchers must be cautious to use consistent box size when collecting data for which comparisons are planned (for example, longitudinal studies or cross-cultural studies.)

Limited methodological research has been undertaken on this response scale. In a study on the practice of tipping, Lynn (2002) used a split sample and collected data using both the unbounded write-in scale and a nine-point semantic differential scale. Lynn created summated scales from his survey items and reported that the data collected using "the semantic differential ratings produced a service index with a skewness of -1.60, while the unbounded write-in scale produced a service index with a skewness of .88 " (p. 10). He concluded that use of the unbounded write-in scale provided more normally-distributed data. Note, however that this finding only held true when the items were combined into an index and that the individual items did not appear to be normally distributed (Lynn, personal communication, January 8, 2003).

Because this new response scale appears to offer some intriguing advantages, we desired to investigate, in a small exploratory study, whether the unbounded scale provides reliable, valid measurements of attitudes using social science questions. Specifically, this study sought to address the following questions.

1. When using the unbounded scale, are the responses independent of the size of the response box?
2. Do items utilizing the unbounded write-in scale exhibit good item test-retest reliability?
3. Is the reliability of the item dependent on the level of emotion evoked by the item?
4. Do unbounded write-in scale responses exhibit skew values closer to zero as compared to five-point Likert-type scale responses?
5. How are the responses on the unbounded write-in scale related to the respondents' Likert scale responses?
6. Do unbounded write-in scale responses provide valid measurements of respondent attitudes?

Method

This study was undertaken during the Fall semester of 2002 using as subjects 220 undergraduate students who were participating in research subject pool requirements at a large public institution. Fifty-six percent of the subjects were female and 50 percent were classified as seniors, 23 percent juniors, and the remainder were sophomores, freshmen and graduate students. The students were randomly split into

four groups to examine the item test-retest reliability of a questionnaire using the unbounded write-in response scale. The questionnaire contained ten selected attitudinal items from the 1998 General Social Survey (National Opinion Research Center, 2002). The selected items were chosen because it was believed that they would elicit skewed responses using a Likert-type scale, however, extremely sensitive items that might compromise student participation were not used. The three formatted questionnaires used in the study are shown in the Appendix. Students were requested to take three administrations of the questionnaire at two-week intervals during the month of October: two unbounded write-in questionnaires and an additional questionnaire that utilized a five-point Likert-type scale for all ten items (this scale was the original used on the GSS). Two different unbounded write-in questionnaires were created using the exact same item stems; the only difference between the two questionnaires was in the size of the boxes provided for response. On questionnaire A, the boxes measured .38 inches by .88 inches, and on questionnaire B, the boxes measured .38 inches by 1.76 inches. In essence, the length of the box for questionnaire B was twice that of questionnaire A. Students first were split into two groups – 110 students were to take the unbounded "short box" questionnaire (questionnaire A) twice and the Likert-type questionnaire (questionnaire C) once and the other 110 students were to take the unbounded "long box" questionnaire (questionnaire B) twice and the Likert-type questionnaire (questionnaire C) once. To control for questionnaire order effects, each of the initial groups were further subdivided into two 55-person groups. One group would receive the Likert-type questionnaire first and then take the two unbounded scale questionnaires on their 2nd and 3rd administrations and the other group would receive the unbounded scale questionnaires first and receive the Likert-type questionnaire on their 3rd administration. Thus the order of the surveys were as follows: Group 1: A A C; Group 2: C A A; Group 3: B B C; Group 4: C B B.

In addition to the questionnaire administration, select students were requested to participate in focus groups after the third administration of the questionnaire. Seven focus groups were constructed: six of the groups were homogeneous with regard to their responses on a specific question (three groups were formed from students responding D, DD, and DDD to the item "Most men are better suited emotionally for politics than are most women" and three groups were formed from students responding A, AA, and AAA to the item "A law which would require a person to obtain a police permit before he or she could buy a gun"). The remaining focus group was constructed of students who tended to report either neutral or were conservative in their use of As and Ds across all 10 items on the questionnaire. All focus groups included five to seven participants and lasted about one hour.

The focus group protocol consisted of two parts. First, participants were asked to read a brief vignette that described a fictional character. They were then asked to respond to the respective survey item for which the group was selected as if they were the character depicted in the scenario. For example, the vignette read by the groups selected for their responses to the item "Most men are better suited emotionally for politics than are most women" included such statements as:

> Pat had volunteered on the campaign of a female candidate, Judy Smith, for state representative…Pat saw that it sometimes was difficult for Judy to hold back her emotions when debating about family issues and addressing personal attacks in the media…"

The vignettes were designed to gauge whether participants responded similarly when exposed to the same information and to elicit discussion about the factors considered when responding using the unbounded scale. Second, groups were asked a series of questions about how their beliefs, attitudes and experiences influenced their responses to both the likert-type and unbounded scale. Because of missed appointments, not all of the 220 students participated in all three questionnaire administrations, however, it appeared from conversations with these students that observations should be missing completely at random. Responses from a total of 190 students are used in the following analyses. Six students participated in only the first administration and an additional eight students were able only to participate in the first and second administrations. These fourteen students with some missing data are

included in the analyses for which they have the appropriate data. All statistical analyses were accomplished using SAS software (Version 8.02) aside from the multilevel analysis described below. An alpha level of .05 was used for testing all hypotheses. Data for the Likert-scale responses were recoded such that "strongly disagree/oppose" received a value of -2, "disagree/oppose" received a value of -1, "neutral" received a value of 0, "agree/approve" received a value of 1, and "strongly agree/approve" received a value of 2. Likewise, data for the unbounded write-in scale were recorded such that each "D" or "O" received a -1 and each "A" received a +1. So, for example, the response "DDD" would be coded as -3 and the response "AA" would be coded as 2.

To answer the first research question, analysis of variance (ANOVA) was undertaken for each of the ten items across the two groups (unbounded short box and unbounded long box) and an accompanying Levene's test for homogeneity of variance was undertaken. For the ANOVAs, only the responses from the first administration of the unbounded write-in scale were used. Because the results suggested that the responses are independent of the size of the boxes (as will be discussed in the results section), observations for the two groups were combined as one "unbounded write-in scale" group and were used in all subsequent analyses.

To address whether the response scale exhibited good test-retest reliability, two different methods were used: Pearson correlations were calculated between the first and second administrations of the unbounded write-in scale, additionally, in order to more concisely summarize the information, a multilevel analysis was undertaken using HLM software (Bryk, Raudenbush, and Congdon, 1996). Prior to the multilevel analysis, the item responses were standardized for each item. The items were treated as the second level of analysis and respondents were considered level one units, clustered within the items. It should be noted, however, that people are not unique to the items and are totally cross-classified. This may result in underestimated standard errors for the item parameters and this violation of the assumption of uncorrelated residuals will be addressed further in the results section. Briefly, multilevel analysis allows the analyst to parse the residual variance at two levels – the individual level (respondent) and the item level. Item test-retest correlation coefficients can be conceptualized as a standardized regression coefficient of the second response regressed on the first response with no intercept in the model.

$$\mathbf{Y_{ij}} = \beta_{1j}\mathbf{X_{ij}} + \mathbf{r_{ij}}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Where Y_{ij} represents the response from the second administration of the unbounded write-in scale for person i on item j, β_{1j} is the slope (or reliability estimate) for item j on the first administration of the item (X_{ij}), and X_{ij} represents the response from the first administration of the unbounded write-in scale for person i on item j. γ_{10} represents the overall slope (or reliability estimate) across the items and u_{1j} represents the residual for item j from that overall slope. Note that if the variance of u_{1j} is significantly different from zero, then it can be concluded that the reliability of the item differed depending on the item. The value of γ_{10} provides an overall reliability coefficient for the ten items and the variance of u_{1j} will inform us whether the regression coefficients significantly vary across the items.

In order to determine if the level of reliability was a function of how emotionally-laden the item was (how far away from neutral the mean responses were), the absolute value of the mean response for the item was entered as a predictor of the item slope, as shown in the following formula:

$$\beta_{1j} = \gamma_{10} + \gamma_{11} ABSMEAN_j + u_{1j}$$

If the coefficient γ_{11} is found to be significantly different from zero and there is significant reduction in the variance of u_{1j}, then it can be concluded that the level of emotion evoked by the item is related to the reliability of the item.

To address the research question regarding the distribution of the responses, the estimates of skew for the Likert-type responses and the first administration of the unbounded write-in scale responses

were compared via t statistics. A descriptive comparison of responses to the Likert-type items and the responses to the unbounded items is provided to address the fifth research question. For example, for the subjects who responded "Strongly Disagree" on the Likert-type questionnaire, their mean response to the first administration of the unbounded scale is provided. Two item stems, those used as the basis for conversation with our focus groups, are investigated in particular.

The final research question was addressed by examining the transcripts from the focus groups. Analysis of focus group data was conducted using techniques typical in qualitative research (Miles and Huberman, 1994; Yin, 1994). First, a matrix was developed to categorize focus group responses by group membership and topic. To determine the contents of each cell, data from focus group transcripts were coded into a priori categories that corresponded with the research questions (e.g., initial perception of the unbounded scale). Cross-group analysis using data from the matrix was utilized to identify overarching themes (Tashakkori & Teddlie, 1998). Transcripts were reviewed to confirm identified themes.

Results

Table 1 displays descriptive statistics on the unbounded responses for the ten items, including the ANOVA results. Surprisingly, as can be seen from the F-statistics, none of the ten analyses suggested that responses (either mean or variability) were dependent on the size of the box. Focus group results support this finding; many of the participants reported that box size did not influence their responses. As one participant stated, "It could have been a blank space, I mean I don't think it really affected me." However, it appears that the existence of the box can affect responses in a variety of ways based on some of the anomalous responses we received. Out of 364 administrations of the unbounded write-in scale items, 8 observations contained atypical responses, including Ds or As written outside of the box and two lines of Ds or As written within the box (both of which are useable observations). Also, we received responses such as "DDD…" and "DDD→" on some of the more emotionally-laden items. In these two cases the dots and arrow filled the remainder of the box. Another anomalous response was "AD," which perhaps was meant to indicate that the person has mixed feelings. None of these latter three situations provide useable data.

Place Table 1 About Here

With regard to reliability, again, the results were somewhat unexpected considering the lack of anchors and bounded endpoints. In general, the unbounded write-in response scale demonstrated modest test-retest reliability. The item correlations appear in Table 2 and ranged from .71 to .86 across the ten items. For use in social group research, while reliability of .7 is seen as modest, a reliability coefficient of .8 is considered sufficient (while reliabilities of at least .9 are desired for individual measurement with high stakes consequences) (Nunnally & Bernstein, 1994). Given these guidelines, the correlations seen here offer some potential. Additionally, the percent of respondents who had an exact match of responses on the first and second administrations of the scale is listed. There is some concern that the percent of respondents with an exact match on the two administrations was somewhat low for some of the items, particularly items 3, 4, and 8. From Table 1, we can see that these three items were also the three most emotionally-laden of the group, perhaps shedding light on the research question about whether the level of emotion evoked by the item would be related to the reliability of the item.

Place Table 2 About Here

The multilevel analysis confirmed and summarized the results from the correlations above. The overall reliability coefficient was .780 ($t=50.42$, $p<.001$), however the variance of u_{1j} was not

significantly different from zero (χ^2=9.94, p=.355) indicating that the reliability coefficients did not vary statistically across the ten items (although power to identify such variance is extremely low with only ten items). Because the variance in the slope residuals was not significantly different from zero, adding the absolute value of the item mean into the equation for the reliability estimate did not decrease the variance of u_{1j} and, in fact, the model fit worsened (as evidenced by a higher deviance value). In future research, to examine differences in reliability across items, a questionnaire with many more items would need to be used.

The skewness of the responses from the first administration of the unbounded write-in scale and the responses from the Likert-type scale are shown in Table 3. T-statistics were calculated to compare these estimates, but because these are related samples, these t-statistics should be viewed as somewhat conservative.

Place Table 3 About Here

Although four out of the ten items exhibited statistically significantly different skew values across the two types of response scales, there was no trend with regard to which response scale provided skew values closer to zero. In addition, the skew values themselves were quite modest and were not large enough such that the analyst would need to use alternate statistical techniques; most traditional techniques are robust to modest violations of normality assumptions (such as the skew values resulting from this analysis).

Table 4 contains information that allows us to look at the distribution of responses on the first administration of the unbounded write-in scale as compared with the same subjects' responses to the Likert-type scale. Because each subject completed both questionnaires, we can directly compare the responses. Once coded, the Likert responses range from -2 to 2 and the unbounded scale responses demonstrated greater variability for each item. However, there is a clear trend that the item means, based on the unbounded write-in scale, correspond to each of the Likert-type responses. For example, for item 3, "Most men are better suited emotionally for politics than are most women," subjects who responded "strongly disagree" on the Likert-type scale averaged -3.16 points on the unbounded scale, while subjects who responded "disagree" averaged just -1.70 on the unbounded scale. Neutral respondents on the Likert-type scale to this item had an average score of -0.38 on the unbounded scale and subjects who responded "agree" had a 0.66 score on average on the unbounded scale. Fewer than five subjects responded "strongly agree" to this statement. While the mean responses on the unbounded scale correspond well to the Likert-type responses across all ten items, it should be noted that there is variability within each Likert-type category. As examples, the data for the two items that were studied in our focus groups, items 3 and 8, are shown in Figures 2 and 3. It is quite interesting that of the 24 respondents who reported "neutral" when using the Likert-type scale for item 3, just five reported "N" on the unbounded scale; ten of these subjects reported "D" when using the unbounded scale (Figure 2). More troubling is the great amount of overlap that can be seen in the "Approve" and "Strongly Approve" categories for item 8 (Figure 3). While the two categories have distinct means on the unbounded scale (2.02 and 2.96, respectively), their distributions are very similar. These findings suggest that either the Likert-type scale is creating some amount of measurement error, with respondents not able to discriminate their feelings between "approve" and "strongly approve," or that the unbounded scale is being interpreted in different ways by these two groups.

Results from the focus groups provided some insight into how participants perceived the two scales, including information on how participants interpreted the unbounded scale. Most focus group participants had never encountered an unbounded scale before participating in this study. When asked about initial reactions to Marder's unbounded write-in scale, participants commonly responded that it elicited more thoughtful responses; participants reportedly deliberated longer when deciding how many

letters to include in their response to the unbounded scale than when choosing between the five anchor points on the Likert-type scale. For example, one participant said that she had "…never answered a question like that or been asked to say how much I felt about something. Usually I know what I like and don't like, but I'd never thought to that degree." Perhaps this practice of carefully contemplating answers is linked to the fact that most focus groups participants believed the unbounded scale represented their opinions and attitudes more accurately than Likert-type scales. When asked whether the unbounded scale was a good measure of attitudes, one participant responded, "you can really tell if people feel really strongly." Interestingly, despite the perceived accuracy of the unbounded scale in reflecting opinions, participants overwhelmingly preferred the Likert-type scale over the Marder scale. Comments suggested that this was perhaps because they thought the unbounded scale required more effort and self-reflection. When asked why he preferred one scale over the other, one participant answered, "I think it's a more difficult way to do it…I think I changed my answer on every single survey…three A's were kind of like, well, what does that mean? It could mean something totally different for each of us. My one A could mean his two A's." So while some focus group members thought that the unbounded scale was able to represent their personal feelings, they were hesitant to endorse using the scale to compare two people's answers. While several participants believed three letters on the unbounded scale represented *strongly agree* on the Likert-type scale and that two letters represented *agree* (findings that are consistent with the data in Figures 2 and 3), it became evident through focus group discussions that there is not a consistent interpretation of what the unbounded scale responses represent. First, while some thought the language in the unbounded scale's introductory statement represented the maximum expected response—the current language uses examples of L, LL, and LLL—others felt the introductory language was ambiguous. As one participant shared, "they don't tell you three A's means this and two A's means that. Everyone can have a different definition of what however many A's or O's means." There was a common belief among focus group members that responses to the unbounded scale would be difficult to interpret because the meaning behind responses was subjective and could vary among respondents. As one focus group participant reflected,

> "You know, you're going to check a box and so is everyone else, but I may write 10 A's on my paper and someone else may just write one. But who's to say that we both didn't agree the same amount? It's all defined by what my scale is. Maybe ten [A's] isn't very much for me."

In addition, it appears that the unbounded scale may be more susceptible to changes in mood or attitude than other scales. Participants frequently reported that their response (i.e., the number of letters written) would depend on their experiences and feelings on the day they completed the survey. The lack of consistent interpretation is evident in statements like "I could've put 12 D's, it wouldn't have mattered, I still disagree. I just picked two [D's]." or comments such as "It's all relative though, to that day or to the person…I mean her three [A's] could be my two [A's]. It depends on what you're representing that A as." The overlap in responses seen in Figure 3 suggested that the unbounded scale may have been interpreted in different ways by two groups—a concern that was articulated during the focus groups. Simply stated by a focus group member, the potential problem is that "two people can feel just the same about a certain topic and one person could put 10 [A's] down and one could be three and they're both thinking the same thing."

Discussion

In general, the results suggest that this response scale holds some promise. However, the main reason for using it, to obtain responses that are more normally distributed than Likert-type scale responses, was not able to be examined. Our Likert responses did not exhibit extreme skew. The reliability of the scale responses was fairly impressive, although coefficients above .8 or closer to .9 would be preferred.

An area of concern, however, is evidenced by some of the statements from the focus group participants. There was some consensus that the responses cannot be compared between survey respondents because of the lack of response scale anchors. However, paradoxically, many of the focus group participants believed that the unbounded scale better reflected their opinions. So, while the unbounded response scale exhibited acceptable reliability, discussions with focus group members have highlighted the doubt that this response scale can be used for comparative research and between-person statistical modeling. An area not examined in this manuscript, however, is the possibility that this scale may provide for adequate within-person measurement, especially for the purposes of ranking items, issues, or statements.

It should be noted that this study has some limitations. First, the questionnaire was a laboratory situation, and thus is likely to suffer from unrealistic conditions. The subjects, college students, were aware that they were in a study and therefore may not have responded in a way that reflected their true beliefs. From focus groups discussions, it is clear that some subjects tried to think back to how they answered the questionnaire the first time and this likely has resulted in an inflated reliability estimate than would be seen in a realistic survey administration. One person's statement, "I tried to remember, [but] I couldn't remember. I was trying to put the same thing," was echoed by several in the focus groups. Yet, interestingly, when participants were asked to recall their answer to either question 3 and 8, few were able to accurately recall their response. A way to avoid this laboratory problem might be to embed just a few questions for test-retest purposes within two larger surveys, for which the majority of items are not given twice. In addition, a two-week interval between administrations might be too short and future research should consider using longer breaks between administrations. A further limitation to the study was a result of outside events. Item 8 on the questionnaire addressed the issue of gun control and from the first to the last administration of the questionnaire, the Washington, D.C. sniper situation unfolded and was resolved. This line of events may have altered students' opinions of the desirability of gun control laws and may thus have negatively affected the test-retest reliability coefficient and our ability to compare responses across administrations.

Summary

The collection of attitudinal data is fairly problematic and prone to measurement error. The more reliable and valid the response scale, the less error prone our measurements will be. While touted for its ability to obtain normally-distributed data, the unbounded write-in scale may at first glance offer questionable reliability and validity. Our findings suggest that the unbounded write-in scale may offer a reliable alternative to the Likert-type scale, although the advantages of its distributional qualities were not seen in this study. Focus group comments, however, lead us to believe that the scale might not reflect similar attitudes across individuals. We suggest additional research directly comparing Likert-type and unbounded response scale reliability and, additionally, a more detailed analysis of how the scale is interpreted by respondents. It is hoped that this study offers some indications to social science researchers of the value of this response scale.

References

Alwin, D. F. (1992). Information transmission in the survey interview: Number of response categories and the reliability of attitude measurement. *Sociological Methodology, 22*, 83-118.

Alwin, D. F., & Krosnick, J. A., (1991). The reliability of survey attitude measurement. *Sociological Methods and Research, 20*, 139-181.

Bryk, A., Raudenbush, S. & Congdon, R. (1996). HLM: Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs (Version 5.4) [Computer software]. Chicago, IL: Scientific Software International, Inc.

Cox, E. P. (1980). The optimal number of response alternative for a scale: A review. *Journal of Marketing Research, 17*, 407-422.

Deshpande, J. V., Gore, A. P., & Shanubhogue, A. (1995). *Statistical Analysis of Nonnormal Data*. San Francisco: Jossey-Bass.

Fouladi, R.T. (2000) Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate nonnormality. *Structural Equation Modeling, 7*, 356-410.

Jacoby, J., & Matell, M. S. (1971). Three-point Likert scales are good enough. *Journal of Marketing Research, 8*, 495-500.

Lynn, M. (2002). Restaurant tips and service quality: A weak relationship or just weak measurement? Unpublished manuscript.

Marder, E. (1997). *The Laws of Choice: Predicting Consumer Behavior*. New York: The Free Press.

Masters, J. R. (1974). The relationship between number of response categories and reliability of likert-type questionnaires. *Journal of Educational Measurement, 11*, 49-53.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.

National Opinion Research Center. Retrieved July 8, 2002 from http://www.icpsr.umich.edu/GSS/.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory*. New York: McGraw-Hill, Inc.

Parill, S. (1999). *Revisiting rating format research: computer-based rating formats and components of accuracy*. Unpublished master's thesis. Virginia Polytechnic Institute, Blacksburg, Virginia.

Schuman, H., & Presser, S. (1981). *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context*. New York: Academic Press.

Sheatsley, P. B. (1983). Questionnaire construction and item writing. In Rossi, P. H., Wright, J. D., & Anderson, A. B. (Eds.) *Handbook of Survey Research*. New York: Academic Press.

Sudman, S., & Bradburn, N. M. (1983). *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey-Bass.

Tashakkori, A., & Teddlie, C. (1998). *Mixed Methodology: Combining qualitative and quantitative approaches*. Thousand Oaks: Sage.

Yin, R. (1994). *Case study research: Design and methods*. (2nd ed.). Thousand Oaks: Sage.

Table 1
Descriptive information for first administration of unbounded write-in scale, by box size

	Short Box			Long Box					
	n	Mean	SD	n	Mean	SD	ANOVA-F	P	Levene-F	p
Item 1	95	0.32	2.22	89	0.34	2.17	0.00	.95	0.03	.86
Item 2	94	-1.07	1.67	89	-1.07	1.73	0.00	.98	0.08	.78
Item 3	95	-1.20	2.09	89	-1.55	2.14	1.26	.26	0.02	.88
Item 4	94	-2.57	2.27	87	-2.97	2.32	1.31	.25	0.02	.89
Item 5	95	1.16	1.85	89	1.26	1.72	0.15	.70	0.23	.63
Item 6	95	1.44	1.85	88	1.52	1.99	0.08	.78	0.22	.64
Item 7	95	1.09	2.02	88	0.67	2.41	1.68	.20	0.90	.35
Item 8	95	1.72	2.38	89	1.58	2.59	0.13	.72	0.27	.60
Item 9	95	1.17	2.31	88	1.10	1.85	0.05	.83	1.47	.23
Item 10	95	-1.09	1.70	89	-1.17	1.50	0.10	.76	0.48	.49

Table 2
Pearson correlations and percent with exact match of first and second unbounded write-in scale responses (short and long box combined)

	n	r	p	exact match
Item 1	179	.86	<.01	50.3%
Item 2	179	.71	<.01	51.4%
Item 3	179	.78	<.01	38.0%
Item 4	177	.80	<.01	35.0%
Item 5	179	.76	<.01	49.7%
Item 6	178	.75	<.01	50.6%
Item 7	178	.73	<.01	53.9%
Item 8	180	.79	<.01	42.8%
Item 9	178	.79	<.01	57.9%
Item 10	179	.84	<.01	63.1%

Table 3
A comparison of skew values for the unbounded and Likert-type responses

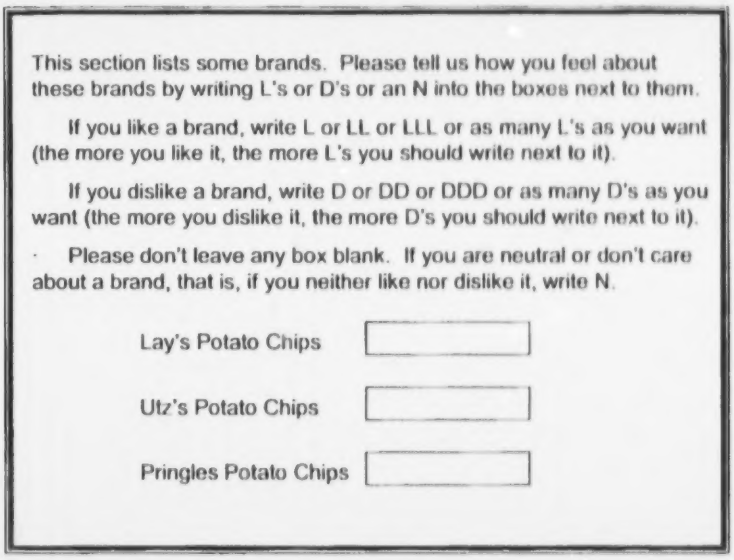
	Unbounded		Likert			
	n	Skew	n	Skew	t	p
Item 1	184	-.526	186	-.194	-1.31	.19
Item 2	183	.133	185	.599	-1.84	.07
Item 3	184	-.600	185	.386	-3.89	<.01
Item 4	181	-.828	185	1.328	-8.51	<.01
Item 5	184	-1.116	186	-.962	-0.61	.54
Item 6	183	.307	186	-.973	5.05	<.01
Item 7	183	-.868	186	-.564	-1.20	.23
Item 8	184	.648	186	-.953	6.32	<.01
Item 9	183	.823	186	-.421	4.91	<.01
Item 10	184	.021	186	.337	-1.25	.21

Table 4
Item means, minimum and maximums for the first administration of the unbounded write-in scale and mean unbounded responses for each Likert-type scale response

	Mean	Min	Max	SD	D	N	A	SA
Item 1	0.34	-8	7	-2.95	-1.12	0.08	1.61	3.35
Item 2	-1.07	-7	3	-2.26	-1.49	-0.53	0.92	---
Item 3	-1.38	-9	3	-3.16	-1.70	-0.38	0.66	---
Item 4	-2.79	-11	3	-4.06	-2.00	-1.08	---	---
Item 5	1.20	-6	6	---	-0.74	0.20	1.54	2.52
Item 6	1.46	-4	10	---	-0.94	0.43	1.41	3.02
Item 7	0.89	-11	8	-3.91	-0.97	0.34	1.62	3.16
Item 8	1.67	-4	13	---	-1.04	0.40	2.02	2.96
Item 9	1.13	-5	12	---	-1.00	0.23	1.73	3.04
Item 10	-1.12	-7	7	-2.55	-1.47	-0.29	0.43	---

--- cell has fewer than 5 observations

Figure 1
Example item using Marder's unbounded write-in scale


This section lists some brands. Please tell us how you feel about these brands by writing L's or D's or an N into the boxes next to them.
If you like a brand, write L or LL or LLL or as many L's as you want (the more you like it, the more L's you should write next to it).
If you dislike a brand, write D or DD or DDD or as many D's as you want (the more you dislike it, the more D's you should write next to it).
Please don't leave any box blank. If you are neutral or don't care about a brand, that is, if you neither like nor dislike it, write N.
Lay's Potato Chips
Utz's Potato Chips
Pringles Potato Chips

(Marder, 1997, p. 156)

Figure 2
Distribution of unbounded responses by Likert-type response category:
"Most men are better suited emotionally for politics than are most women"

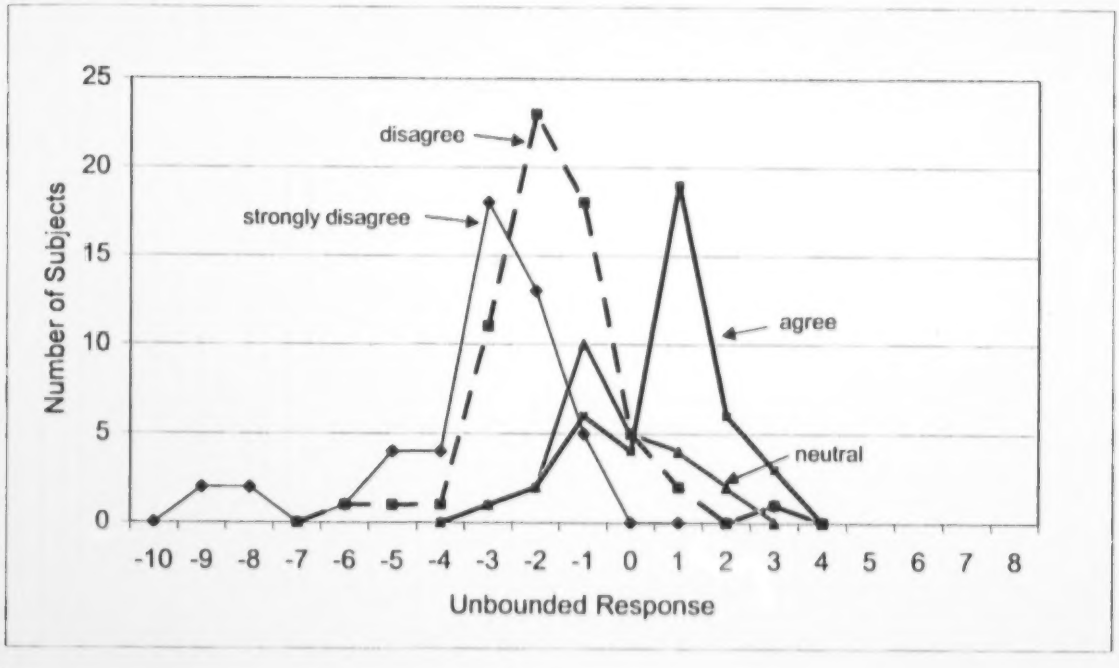
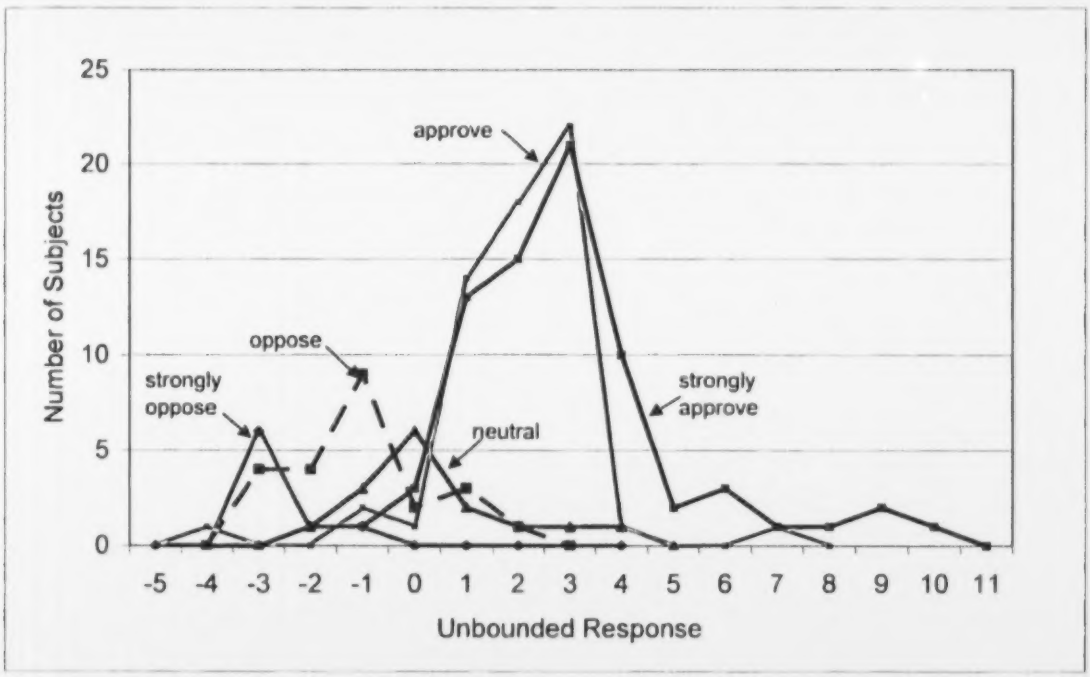


Figure 3
Distribution of unbounded responses by Likert-type response category:
"A law which would require a person to obtain a police permit before he or she could buy a gun"



APPENDIX – Questionnaire A

ID _____

Date _____

A. This section lists some statements. Please tell us the extent to which you agree or disagree with these statements by writing A's or D's or an N into the boxes next to them.

If you agree with a statement, write A or AA or AAA or as many A's as you want (the more strongly you agree, the more A's you should write in the box).

If you disagree with a statement, write D or DD or DDD or as many D's as you want (the more strongly you disagree, the more D's you should write in the box).

Please don't leave any box blank. If you are neutral about a statement, that is, if you neither agree nor disagree, write N in the box.

	1. It is sometimes necessary to discipline a child with a good hard spanking.
	2. Modern painting is just slapped on; a child could do it.
	3. Most men are better suited emotionally for politics than are most women.
	4. Women should take care of running their homes and leave running the country up to men.

Protecting secrets is a continuing concern of the government. The following are security measures that the government might apply to individuals with a SECRET or TOP SECRET clearance. Please tell us the extent to which you agree or disagree (by writing A's and D's or an N) that people with SECRET or TOP SECRET clearance should be subject to:

	5. Periodic lie detector tests.
	6. Random drug tests.

B. This section lists some issues. Please tell us the extent to which you approve of or oppose these issues by writing A's or O's or an N into the boxes next to them.

If you approve of the issue, write A or AA or AAA or as many A's as you want (the more strongly you approve, the more A's you should write in the box).

If you oppose the issue, write O or OO or OOO or as many O's as you want (the more strongly you oppose, the more O's you should write in the box).

Please don't leave any box blank. If you are neutral about the issue, that is, if you neither approve of it nor oppose it, write N in the box.

	7. The death penalty for persons convicted of murder.
	8. A law which would require a person to obtain a police permit before he or she could buy a gun.
	9. Pro-athletes giving thanks to God during sports events.
	10. The use of religious "images" in public advertising to sell non-religious commercial products.

APPENDIX – Questionnaire B

ID _____

Date _____

A. This section lists some statements. Please tell us the extent to which you agree or disagree with these statements by writing A's or D's or an N into the boxes next to them.

If you agree with a statement, write A or AA or AAA or as many A's as you want (the more strongly you agree, the more A's you should write in the box).

If you disagree with a statement, write D or DD or DDD or as many D's as you want (the more strongly you disagree, the more D's you should write in the box).

Please don't leave any box blank. If you are neutral about a statement, that is, if you neither agree nor disagree, write N in the box.

	1. It is sometimes necessary to discipline a child with a good hard spanking.
	2. Modern painting is just slapped on; a child could do it.
	3. Most men are better suited emotionally for politics than are most women.
	4. Women should take care of running their homes and leave running the country up to men.

Protecting secrets is a continuing concern of the government. The following are security measures that the government might apply to individuals with a SECRET or TOP SECRET clearance. Please tell us the extent to which you agree or disagree (by writing A's and D's or an N) that people with SECRET or TOP SECRET clearance should be subject to:

	5. Periodic lie detector tests.
	6. Random drug tests.

B. This section lists some issues. Please tell us the extent to which you approve of or oppose these issues by writing A's or O's or an N into the boxes next to them.

If you approve of the issue, write A or AA or AAA or as many A's as you want (the more strongly you approve, the more A's you should write in the box).

If you oppose the issue, write O or OO or OOO or as many O's as you want (the more strongly you oppose, the more O's you should write in the box).

Please don't leave any box blank. If you are neutral about the issue, that is, if you neither approve of it nor oppose it, write N in the box.

	7. The death penalty for persons convicted of murder.
	8. A law which would require a person to obtain a police permit before he or she could buy a gun.
	9. Pro-athletes giving thanks to God during sports events.
	10. The use of religious "images" in public advertising to sell non-religious commercial products.

APPENDIX – Questionnaire C

ID _____

Date _____

A. This section lists some statements. Please tell us the extent to which you agree or disagree with these statements by checking the appropriate box.

B.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. It is sometimes necessary to discipline a child with a good hard spanking.	❒	❒	❒	❒	❒
2. Modern painting is just slapped on; a child could do it.	❒	❒	❒	❒	❒
3. Most men are better suited emotionally for politics than are most women.	❒	❒	❒	❒	❒
4. Women should take care of running their homes and leave running the country up to men.	❒	❒	❒	❒	❒

Protecting secrets is a continuing concern of the government. The following are security measures that the government might apply to individuals with a SECRET or TOP SECRET clearance. Please tell us the extent to which you agree or disagree that people with SECRET or TOP SECRET clearance should be subject to:

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
5. Periodic lie detector tests.	❒	❒	❒	❒	❒
6. Random drug tests.	❒	❒	❒	❒	❒

BEST COPY AVAILABLE

B. This section lists some issues. Please tell us the extent to which you approve of or oppose these issues by checking the appropriate box.

	Strongly Oppose	Oppose	Neutral	Approve	Strongly Approve
7. The death penalty for persons convicted of murder.	❒	❒	❒	❒	❒
8. A law which would require a person to obtain a police permit before he or she could buy a gun.	❒	❒	❒	❒	❒
9. Pro-athletes giving thanks to God during sports events.	❒	❒	❒	❒	❒
10. The use of religious "images" in public advertising to sell non-religious commercial products.	❒	❒	❒	❒	❒

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC Educational Resources Information Center

REPRODUCTION RELEASE

(Specific Document)

TM035226

I. DOCUMENT IDENTIFICATION:

Title: AN EXPLORATION OF THE VALIDITY OF THE UNBOUNDED WRITTEN SCALE	
Author(s): STARLETON, L.M. & EDMONDS, M.	
Corporate Source:	Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1 ↑ ☑

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A ↑ ☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only.

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B ↑ ☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only.

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Laura M. Stapleton	Printed Name/Position/Title: LAURA M. STAPLETON, ASST. PROFESSOR
Organization/Address: UNIVERSITY OF TEXAS, D5800 AUSTIN, TX 78712	Telephone: 512/471-0858 FAX: 512/471-1288
	E-Mail Address: LAURA.STAPLETON@MAIL.UTEXAS.EDU Date: 9/5/03

(Over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20742-5701
ATTN: ACQUISITIONS

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfacility.org

EFF-088 (Rev. 2/2001)